

Next please! A new definition of the treatment and control groups for randomizations with waiting lists.*

Clément de Chaisemartin[†]

Luc Behaghel[‡]

December 1, 2015

Abstract

In many randomized experiments, the treatment is offered to eligible units following a random order, until all available seats are filled. In such instances, the treatment (resp. control) group is often defined as units receiving (resp. not receiving) an offer. In this note, we show that groups constructed this way are not statistically comparable. On the other hand, we show that units with a random number strictly lower than the random number of the last unit which received an offer are statistically comparable to units with a random number strictly greater.

Keywords: Randomized control trials, waiting lists, instrumental variables

JEL Codes: C21, C23

*We are very grateful to Josh Angrist and Xavier D'Haultfoeuille for their helpful comments.

[†]University of Warwick, clement.de-chaisemartin@warwick.ac.uk

[‡]Paris School of Economics - INRA, luc.behaghel@ens.fr

In this note, we consider randomizations with waiting lists. This type of designs has frequently been used in the education literature (see e.g. Abdulkadiroglu et al., 2011, Behaghel et al., 2015, or Curto & Fryer, 2014). To make matters concrete, throughout the note we consider an example where n students have applied to enter into some school where $s < n$ seats are available.

Assume that students are of two types. Some are accepters and will join the school if they receive an offer. Some are refusers and will not join the school even if they receive an offer. Let a_1 and $a_0 = n - a_1$ respectively denote the number of accepters and refusers. Assume also that seats in the school are offered as follows. Each student is assigned a random number R_i included between 1 and n . Students with $R_i \leq s$ receive an offer to join the school. If all of them accept, no other student receives an offer. If r students refuse, students with $s < R_i \leq s + r$ receive an offer, and so on and so forth until the s seats are filled. Let T denote the random number of the last student receiving an offer.

In such designs, a first possible definition of the instrument is $Z_i = 1\{R_i \leq s\}$: students receiving the first round of offers make up the treatment group, while other students make up the control group. Because the number assigned to each student is random, students with $Z_i = 1$ and $Z_i = 0$ are statistically comparable.

Another definition of the instrument that has often been used is $V_i = 1\{R_i \leq T\}$: students receiving an offer make up the treatment group, while other students make up the control group. However, the groups with $V_i = 1$ and $V_i = 0$ might not be statistically comparable. Because the subgroup with $V_i = 1$ bears all students who accepted an offer, one might worry that it bear a greater proportion of accepters than the subgroup with $V_i = 0$, thus making the two subgroups not statistically comparable. For instance, the student with $R_i = T$ must by definition be an accepter (if she had not been an accepter, the school would have needed to offer a seat to the student with rank $R_i = T + 1$ to fill all of its seats).

In what follows, we show that this concern is legitimate, but can be addressed through a very slight modification of V_i . Let $W_i = 1\{R_i < T\} - 1\{R_i = T\}$. The new treatment and control groups we propose are respectively equal to students with $W_i = 1$ and $W_i = 0$. The student with $R_i = T$ has $W_i = -1$ and is excluded from the analysis. The next theorem shows that in expectation across all the possible orderings of students, the subgroups with $W_i = 1$ and $W_i = 0$ bear the same proportion of accepters, thus implying that they are statistically comparable and can be used as valid treatment and control groups.

For any integer i let $i! = i \times (i - 1) \times \dots \times 1$, with the convention that $0! = 1$. For any integers i and j , let $\binom{i}{j} = \frac{i!}{j!(i-j)!}$, with the convention that $\binom{i}{j} = 0$ if $j > i$. Let w_1 and w_0 respectively denote the proportion of accepters in the subset of students with $W_i = 1$ and $W_i = 0$. w_1 and w_0 are random variables: they vary according to the random ordering of students in the waiting list. Our goal is to show that these two random variables have the

same expected value across all possible orderings of students in the waiting list. Before stating and proving the theorem, it might be useful to note the following point. There are $n!$ possible random orderings of the n students in the waiting list. However, we only need to consider the $\binom{n}{a_1}$ possible orderings of accepters and refusers, because w_1 and w_0 do not vary across orderings where accepters and refusers occupy the same positions. To see this point, it might be easier to consider an example. Assume that $n = 6$, $s = 2$, $a_1 = 3$, and students 1, 3, and 5 are accepters. Then, consider the following ordering of students: (2; 3; 5; 4; 1; 6). With this ordering, student 2 first receives an offer but declines it because she is a refuser; then student 3 receives an offer and accepts it because she is an accepter; then student 5 receives an offer and accepts it because she is an accepter; then no other student receives an offer because all seats are taken. Students 2 and 3 have $W_i = 1$, while students 4, 1 and 6 have $W_i = 0$. Therefore $w_1 = 1/2$ and $w_0 = 1/3$. Now, consider another ordering of students: (6; 1; 3; 2; 5; 4). This ordering differs from the previous one, but accepters still come in 2nd, 3rd, and 5th positions in the waiting list. With this ordering we still have $w_1 = 1/2$ and $w_0 = 1/3$.

Theorem 1 For any $(n, s, a_1) \in \mathbb{N}^3 : 2 \leq s < a_1 \leq n$, $E(w_0) = E(w_1) = \frac{a_1}{n}$.

Proof

$$\begin{aligned}
E(w_0) &= \sum_{j=0}^{a_0} E(w_0 | T = s + j) P(T = s + j) \\
&= \frac{1}{\binom{a_1 + a_0}{a_1}} \sum_{j=0}^{a_0} \binom{s-1+j}{s-1} \binom{a_1 + a_0 - (s+j)}{a_1 - s} \frac{a_1 - s}{a_1 + a_0 - (s+j)} \\
&= \frac{1}{\binom{a_1 + a_0}{a_1}} \sum_{j=0}^{a_0} \binom{s-1+j}{s-1} \binom{a_1 + a_0 - 1 - (s+j)}{a_1 - 1 - s} \\
&= \frac{\binom{a_1 - 1 + a_0}{a_1 - 1}}{\binom{a_1 + a_0}{a_1}} \\
&= \frac{a_1}{a_1 + a_0} = \frac{a_1}{n}.
\end{aligned}$$

The first equality follows from the law of iterated expectations, as T must be included between s and $s + a_0$. The second equality follows from the two following facts. First, having $T = s + j$ is equivalent to having $s - 1$ accepters with $R_i \leq s - 1 + j$, one accepter with $R_i = s + j$, and $a_1 - s$ accepters with $R_i \geq s + 1 + j$. Therefore, $P(T = s + j) = \frac{\binom{s-1+j}{s-1} \binom{a_1 + a_0 - (s+j)}{a_1 - s}}{\binom{a_1 + a_0}{a_1}}$. Second, when $T = s + j$, $a_1 - s$ accepters have $W_i = 0$, out of a total of $a_1 + a_0 - (s + j)$ students with $W_i = 0$. Therefore, $w_0 = \frac{a_1 - s}{a_1 + a_0 - (s + j)}$. The fourth equality follows from the following combinatorial reasoning: to compute the number of ways to allocate $a_1 - 1$ students to $a_1 - 1 + a_0$ ranks, one can partition according to the rank of the s th student (under the assumptions of the theorem, $s \leq a_1 - 1$), which can be anywhere between s and $s + a_0$. The number of ways of allocating the s th student to rank $s + j$ is equal to $\binom{s-1+j}{s-1} \binom{a_1 - 1 + a_0 - (s+j)}{a_1 - 1 - s}$.

Similarly,

$$\begin{aligned}
E(w_1) &= \sum_{j=0}^{a_0} E(w_1|T = s+j)P(T = s+j) \\
&= \frac{1}{\binom{a_1+a_0}{a_1}} \sum_{j=0}^{a_0} \binom{s-1+j}{s-1} \binom{a_1+a_0-(s+j)}{a_1-s} \frac{s-1}{s-1+j} \\
&= \frac{1}{\binom{a_1+a_0}{a_1}} \sum_{j=0}^{a_0} \binom{s-2+j}{s-2} \binom{a_1+a_0-(s+j)}{a_1-s} \\
&= \frac{1}{\binom{a_1+a_0}{a_1}} \sum_{j=0}^{a_0} \binom{s-2+j}{s-2} \binom{a_1-1+a_0-(s-1+j)}{a_1-1-(s-1)} \\
&= \frac{\binom{a_1-1+a_0}{a_1-1}}{\binom{a_1+a_0}{a_1}} \\
&= \frac{a_1}{a_1+a_0} = \frac{a_1}{n}.
\end{aligned}$$

The fifth equality follows from a similar combinatorial reasoning as above, except that one now has to partition according to the rank of the $s-1$ th student (under the assumptions of the theorem, $s-1 \geq 1$).

QED.

To illustrate this theorem, we consider a numerical example. Assume that $a_1 = 4$, $n = 6$, and $s = 2$. Table 1 below presents the $\binom{6}{4} = 15$ possible orderings of accepters (A) and refusers (R). As explained above, we do not need to consider the $6! = 720$ possible orderings of individual students, because w_0 and w_1 do not vary across orderings where accepters and refusers have the same positions. For each ordering, students in blue have $W_i = 1$, while those in red have $W_i = 0$. Across the 15 possible orderings,

$$E(w_1) = \frac{1}{15}(1+1+1+1+1+1+1/2+1/2+1/2+1/3+1/2+1/2+1/2+1/3+1/3) = 2/3.$$

Similarly,

$$E(w_0) = \frac{1}{15}(1/2+1/2+1/2+1/2+1/2+1/2+2/3+2/3+2/3+1+2/3+2/3+2/3+1+1) = 2/3.$$

Table 1: $W_i=1$ and $W_i=0$ in a numerical example

A	A	A	A	A	A	A	A	A	A	R	R	R	R	R
A	A	A	A	A	A	R	R	R	R	A	A	A	A	R
A	A	A	R	R	R	A	A	A	R	A	A	A	R	A
A	R	R	A	A	R	A	A	R	A	A	A	R	A	A
R	A	R	A	R	A	A	R	A	A	A	R	A	A	A
R	R	A	R	A	A	R	A	A	A	R	A	A	A	A

Theorem 1 relies on the assumption that $s < a_1$. This assumption is testable. To see this, consider first a simple example where $n = 6$ and $s = 2$. Assume that $T = 4$, meaning that the first four students in the waiting list received an offer. Under the null hypothesis that $a_1 = s = 2$, we can compute the probability to observe $T = 4$. This probability is equal to $\frac{\binom{3}{2}}{\binom{6}{4}} = \frac{1}{5}$. Under the null, 4 of the 6 students are refusers, and there are $\binom{6}{4}$ possible ways of allocating them to the 6 positions in the waiting list. If $T = 4$, then it must be the case that 2 of the first 3 students in the waiting list are refusers, and that the last 2 students are refusers. $\binom{3}{2} \times \binom{2}{2} = \binom{3}{2}$ allocations of the four refusers in the waiting list satisfy this condition, hence the result. Theorem 2 below generalizes this result to arbitrary values of (n, s, T) .

Theorem 2 *For any $(n, s) \in \mathbb{N}^2 : 2 \leq s < n$, under the null hypothesis that $a_1 = s$ the probability of observing that $T = t$ is equal to $\frac{\binom{t-1}{t-s}}{\binom{n}{n-s}}$, for any $t : s \leq t \leq n$.*

Proof

If $a_1 = s$, there are $\binom{n}{n-s}$ possible ways of allocating the $n - s$ refusers to the n positions in the waiting list. Moreover, for any $t : s \leq t \leq n$, to have $T = t$ one must have $t - 1 - (s - 1)$ refusers in the first $t - 1$ positions in the waiting list, and $n - t$ of them in the last $n - t$ positions in the waiting list. $\binom{t-1}{t-1-(s-1)} \times \binom{n-t}{n-t} = \binom{t-1}{t-s}$ allocations of the $n - s$ refusers to the n positions in the waiting list satisfy this condition, hence the result.

QED.

To test the null hypothesis that $a_1 = s$, one merely need to compute $\frac{\binom{T-1}{T-s}}{\binom{n}{n-s}}$ for the actual value of T observed in the data. When this quantity is below some pre-specified level α , one rejects the null. Note that this test does not rely on any asymptotic approximation: it is valid even in a small sample. When $a_1 = s$ is rejected at a sufficiently low level, researchers can confidently use W_i to define their control and treatment groups when analyzing data from a waiting list randomization.

So far, we have considered the case where only one lottery is conducted to allocate treatment. Actually, most papers relying on waiting list randomizations do not use data from a single lottery, but pool together several lotteries. For instance, Behaghel et al. (2015) study the effect of a boarding school for disadvantaged students. To allocate seats, the school conducted 14 lotteries in each gender \times grade cell with more applicants than seats. In their analysis, Behaghel et al. (2015) pool together the data from these 14 lotteries.

Results from this note can still be used when several lotteries are conducted. However, one needs to use inverse probability reweighting to ensure that pooling does not lead again to an imbalance between the share of accepters in the treatment and in the control group. To see this, consider the following example. Assume that one wants to pool data from 15 lotteries conducted in 15 strata which each have 6 students, 4 accepters, and 2 seats. In expectation, the random ordering of accepters and refusers in these 15 strata will be as in Table 1. Indeed,

in each stratum the random ordering of accepters and refusers has a probability $1/15$ of being the same as in column 1 of the table, a probability $1/15$ of being the same as in column 2 of the table, etc. Therefore, in expectation, out of the 15 strata there will be one stratum where the random ordering of accepters and refusers is the same as in column 1, one stratum where the ordering is the same as in column 2, etc. Pooling naively the students with $W_i = 1$ across the 15 strata yields in expectation a treatment group made up of 27 students, 15 of which are accepters. On the other hand, pooling the students with $W_i = 0$ across the 15 strata yields in expectation a control group made up of 48 students, 30 of which are accepters. In expectation, $5/9$ of students in the treatment group are accepters, against $5/8$ in the control group. This example shows that while within each stratum our definition of the treatment and control groups produces statistically identical groups, pooling naively lottery strata will again create an imbalance. This imbalance can easily be corrected, for instance through inverse probability reweighting (see e.g. Rosenbaum & Rubin, 1983). Students with $W_i = 1$ should receive a weight equal to the ratio of the share of students with $W_i = 1$ in the population and in their stratum. On the other hand, students with $W_i = 0$ should receive a weight equal to the ratio of the share of students with $W_i = 0$ in the population and in their stratum. In the example considered in Table 1, students with $W_i = 1$ in the first stratum (column 1 of the table) should receive a weight equal to $\frac{27/75}{1/5}$, while students with $W_i = 0$ should receive a weight equal to $\frac{48/75}{4/5}$. In the reweighted sample, the share of accepters among students with $W_i = 1$ is

$$\frac{6 \times 5 + 6 \times 5/2 + 3 \times 5/3}{6 \times 5 + 12 \times 5/2 + 9 \times 5/3} = 2/3.$$

Similarly, the share of accepters among students with $W_i = 0$ is

$$\frac{12 \times 5/4 + 12 \times 5/3 + 6 \times 5/2}{24 \times 5/4 + 18 \times 5/3 + 6 \times 5/2} = 2/3.$$

Therefore, reweighting reestablishes balance between the treatment and the control group.¹ On the other hand, including strata fixed effects in the regression (the standard practice in most papers using waiting list designs and pooling data from several lotteries) is not sufficient to restore balance between the treatment and the control group, as we illustrate in the Monte-Carlo study below.

Let $\hat{\beta}_W$, $\hat{\beta}_V$, and $\hat{\beta}_Z$ denote the coefficient of treatment in a 2SLS regression of some outcome Y_i on some treatment D_i using W_i , V_i , and Z_i as the instrument. It follows from Theorem 1 that $\hat{\beta}_V$ is not a consistent estimator of the average treatment effect among accepters. On the other hand, both $\hat{\beta}_W$ and $\hat{\beta}_Z$ are consistent. But using standard results on 2SLS regression, it is easy to show that when n goes to infinity, the ratio of the asymptotic variances of $\hat{\beta}_W$ and $\hat{\beta}_Z$ is equal to

$$\frac{p_c - p_d}{1 - p_d}, \tag{1}$$

¹This result is not specific to the problem we consider here. It was already proven in a number of different contexts (see Rosenbaum & Rubin, 1983 or Frölich, 2007).

where p_c and p_d respectively denote the share of accepters in the population, and the share of treated units in the population. This ratio is always lower than 1, because using W_i instead of Z_i always increases the take-up difference between the treatment and the control group. For instance, for $p_c = 0.8$ and $p_d = 0.6$ this ratio is equal to $1/2$. Therefore, using W_i instead of Z_i can lead to important gains in statistical precision.

We illustrate these results with a small Monte-Carlo study. We consider a DGP with 20 lottery strata. In each strata, there are 10 seats for treatment and 20 students, 15 of which are accepters. Y_{0i} , the outcome of students without the treatment, follows a $N(0, 1)$ distribution for refusers and a $N(1, 1)$ distribution for accepters. Y_{1i} , their outcome with the treatment, is equal to $Y_{0i} + 0.2$ for accepters, and to Y_{0i} for refusers. This implies that every accepter in the population has a treatment effect of 0.2. We allocate to each student a random number R_i included between 1 and 20 and representing her rank in the waiting list of her strata. Accordingly, all students with a lottery number below that of the 10th accepter in their stratum receive a treatment offer, but only the 10 accepters actually get treated. We also use R_i to construct the three instruments considered in this note, W_i , V_i , and Z_i . Finally, we estimate 2SLS regressions of Y_i on D_i using either V_i , W_i , or Z_i as the instrument, and using either strata fixed effects or inverse probability reweighting to pool the data from the 20 strata.² We replicate 2000 times this procedure, and report the average, the standard error, and the confidence intervals of the five 2SLS coefficients in Table 2 below. As expected, using V_i as the instrument and strata fixed effects does not yield a consistent estimator of the average treatment among accepters, as reflected by the fact that 0.2 does not lie within its confidence interval. On average, the resulting 2SLS coefficient is 16% greater than the target parameter. The magnitude of the bias depends on the extent to which the distribution Y_{0i} differs across refusers and accepters: in a DGP where Y_{0i} follows a $N(0, 1)$ distribution among refusers and a $N(0.5, 1)$ distribution among accepters, the bias is smaller. Using V_i as the instrument and reweighting reduces the bias, but the resulting 2SLS coefficient is still inconsistent. On average it is 11% greater than the target parameter. Using W_i and strata fixed effects further reduces the bias, but the resulting 2SLS coefficient is still inconsistent. On average, it is 6% greater than the target parameter. On the other hand, using W_i and reweighting yields a consistent estimator. Using Z_i also yields a consistent estimator, but the standard error of the corresponding 2SLS coefficient is substantially larger than that obtained using W_i and reweighting. The ratio of the standard errors of $\hat{\beta}_W$ and $\hat{\beta}_Z$ is equal to 0.77, which is close to what we predicted in Equation (1) using an asymptotic approximation: $\sqrt{\frac{0.75-0.5}{1-0.5}} \approx 0.71$.

²When the instrument is Z_i , using reweighting or strata fixed effects yields the same 2SLS coefficient because the share of units with $Z_i = 1$ is the same in all strata.

Table 2: Results of a small Monte-Carlo study

	Average	SE	95% CI
Instrument V, fixed effects	0.2311	0.1313	[0.2254,0.2369]
Instrument V, reweighting	0.2203	0.1319	[0.2145,0.2261]
Instrument W, fixed effects	0.2111	0.1369	[0.2051,0.2171]
Instrument W, reweighting	0.2014	0.1376	[0.1954,0.2074]
Instrument Z	0.1982	0.1792	[0.1904,0.2061]

References

- Abdulkadiroglu, A., Angrist, J., Dynarski, S. M., Kane, T. J. & Pathak, P. A. (2011), ‘Accountability and flexibility in public schools: Evidence from boston’s charters and pilots’, *The Quarterly Journal of Economics* **126**(2), 699–748.
- Behaghel, L., Chaisemartin, C. d. & Gurgand, M. (2015), ‘Ready for boarding?: the effects of a boarding school for disadvantaged students’.
- Curto, V. E. & Fryer, R. G. (2014), ‘The potential of urban boarding schools for the poor: Evidence from seed’, *Journal of Labor Economics* **32**(1), 65–93.
- Frölich, M. (2007), ‘Nonparametric iv estimation of local average treatment effects with covariates’, *Journal of Econometrics* **139**(1), 35–75.
- Rosenbaum, P. R. & Rubin, D. B. (1983), ‘The central role of the propensity score in observational studies for causal effects’, *Biometrika* **70**(1), 41–55.